Probability Distribution
Lecture 10

Lecture objectives

- Find the moment and moment generating function for a random variable.
- Find the exact probability for X successes in n trials of a binomial experiment.
- Find the mean, variance, and standard deviation for the variable of a binomial distribution.

1 The Moments

The *nth* moment of a distribution (or set of data) about a number is the expected value of the *nth* power of the deviations about that number. In statistics, moments are needed about the mean, and about the origin.

- The *nth* moment of a distribution about zero is given by $E(X^n)$
- The *nth* moment of a distribution about the mean is given by $E(X-\mu)^n$

Then each type of measure includes a moment definition. For example The expected value, E(X), is the first moment about zero and the variance, Var(X), is the second moment about the mean, $E(X-\mu)^2$.
The rule of *nth* moment about zero is:

$$M_0^n = E(X^n) = \sum x^n P(x)$$

Or

$$M_0^n = E(X^n) = \int x^n f(x)\,dx$$

1.1 Moment Generating Functions

Moment generating functions are function of t $M_X(t) = E\left[e^{tX}\right]$. provided that this expectation for a value of t. Moment generating functions are useful for generating the moments of X to compute the *nth* moment of X. The moment generating function of X is

$$M_X(t) = E\left[e^{tX}\right] = E\left[\exp(tX)\right]$$
$$M_X(t) = E(e^{tX}) = \sum_{\text{all } x} e^{tx} P(x)$$

or

$$M_X(t) = E(e^{tX}) = \int_x e^{tx} f(x)\,\mathrm{d}x$$

Note that exp(X) is another way of writing e^X.
When the nth derivative (with respect to t) of the moment generating function is evaluated at t=0, the nth moment of the random variable X about zero will be obtained.

$$E(X^n) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$$

Example 1. Suppose a discrete PDF is given by the following table

X=x	P(x)
0	0.4
1	0.35
2	0.25

obtain the moment generating function and obtain the moments about zero.

Solution

$$\begin{aligned} M_X(t) &= E(e^{tx}) \\ &= \sum_{\text{all } x} e^{tx} P(x) \\ &= 0.4e^{0t} + 0.35e^{1t} + 0.25e^{2t} \\ &= 0.4 + 0.35e^{t} + 0.25e^{2t} \end{aligned}$$

We can then compute derivatives and obtain the moments about zero.

$$\begin{aligned} M'_X(t) &= 0.35e^{t} + 0.5e^{2t} & M'_X(0) &= 0.35 + 0.5 = 0.85 \\ M''_X(t) &= 0.35e^{t} + e^{2t} & M''_X(0) &= 0.35 + 1 = 1.35 \\ M^{(3)}_X(t) &= 0.35e^{t} + 2e^{2t} & M^{(3)}_X(0) &= 0.35 + 2 = 2.35 \end{aligned}$$

then

$$\begin{aligned} E(X) &= M'(0) = 0.85 \\ Var(X) &= E(X^2) - E(X)^2 \\ Var(X) &= 1.35 - 0.85^2 = 0.6275 \end{aligned}$$

Example 2. Let us now consider the pdf given by $f(x) = 4e^{-4x}$ defined on the interval $[0, \infty]$. Find the moment generating function

Solution

$$\begin{aligned} M_X(t) &= E(e^{tx}) \\ &= \int_0^{\infty} e^{tx} f(x)\,\mathrm{d}x \\ &= \int_0^{\infty} e^{tx} 4e^{-4x}\,\mathrm{d}x \\ &= \int_0^{\infty} 4e^{(t-4)x}\,\mathrm{d}x \\ &= \frac{4}{t-4} e^{(t-4)x}\Big|_0^{\infty} \\ &= \frac{-4}{t-4} = -4(t-4)^{-1} \end{aligned}$$

2 Discrete Probability Distributions

2.1 Binomial Distribution

A binomial experiment is a probability experiment that satisfies the following four requirements:

1. There must be a fixed number of trials.
2. Each trial can have only two outcomes or outcomes that can be reduced to two outcomes. These outcomes can be considered as either success or failure.
3. The outcomes of each trial must be independent of one another.
4. The probability of a success must remain the same for each trial.

The outcomes of a binomial experiment and the corresponding probabilities of these outcomes are called a *binomial distribution.*
In a binomial experiment, the probability of exactly X successes in n trials is
$P(X) =_n C_X.p^X.q^{(n-X)}$
Where
p=The numerical probability of a success
q= The numerical probability of a failure
$1-p=q$
n=The number of trials
X=The number of successes in n trials
Note that $0 \leq X \leq n$ and $X = 0, 1, 2, 3, ..., n.$

Example 3. A coin is tossed 3 times. Find the probability of getting exactly two heads.

Solution

looking at the sample space. There are three ways to get two heads.
HHH, HHT, HTH, THH, TTH, THT, HTT, TTT
The answer is$\frac{3}{8}$, or 0.375.
it meets the four requirements of the binomial expiermint
In this case, n=3, X=2, p=0.5 , and q=0.5. Hence, substituting in the formula gives $P(2heads) = P(X = 2) =_n C_X.p^X.q^{(n-X)} =_3 C_2.(o.5^2).(0.5^{(3-2)}) =$ 0.375

2.1.1 The mean, variance, and standard deviation

The mean, variance, and standard deviation of a variable that has the binomial distribution can be found by using the following formulas.
$Mean : \mu = n.p \; Variance : \sigma^2 = n.p.q \; Standarddeviation : \sigma = \sqrt{n.p.q}$

Example 4. A coin is tossed 4 times. Find the mean, variance, and standard deviation of the number of heads that will be obtained.

Solution

With the formulas for the binomial distribution and n=4, p=0.5 , and q=0.5 , the results are
$\mu = n.p = 4 * 0.5 = 2$
$sigma^2 = n.p.q = 4 * 0.5 * 0.5 = 1$
$\sigma = 1$
solve this problem by using the formulas for expected value and The probality distribution.

2.2 The Poisson Distribution

A discrete probability distribution that is useful when n is large and p is small and when the independent variables occur over a period of time is called the Poisson distribution.
The Poisson distribution can be used when a density of items is distributed over a given area or volume, such as the number of plants growing per acre or the number of defects in a given length of videotape.
The probability of X occurrences in an interval of time, volume, area, etc., for a variable where λ (Greek letter lambda) is the mean number of occurrences per unit (time, volume, area, etc.) is

$$P(X;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Where X=1,2,3,... and the letter **e** is a constant approximately equal to 2.7183

Example 5. If there are 200 typographical errors randomly distributed in a 500-page manuscript, find the probability that a given page contains exactly 3 errors.

Solution

First, find the mean number l of errors. Since there are 200 errors distributed over 500 pages, each page has an average of
$\lambda = \frac{200}{500} = \frac{2}{5} = 0.4$
or 0.4 error per page. Since X =3, substituting into the formula yields

$$P(X;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{2.7183^{-0.4}0.4^3}{3!} = 0.0072$$

Example 6. A sales firm receives, on average, 3 calls per hour on its toll-free number. For any given hour, find the probability that it will receive the following.

1. At most 3 calls
2. At least 3 calls

Solution

1. "At most 3 calls" means 0, 1, 2, or 3 calls. Hence,
   P(0; 3) + P(1; 3) + P(2; 3) + P(3; 3)=
   0.0498 + 0.1494 + 0.2240 + 0.2240= 0.6472

2. "At least 3 calls" means 3 or more calls. It is easier to find the probability of 0, 1, and 2 calls and then subtract this answer from 1 to get the probability of at least 3 calls.
   P(0; 3) + P(1; 3) + P(2; 3) = 0.0498 + 0.1494 + 0.2240 = 0.4232
   and
   1 - 0.4232 = 0.5768

2.2.1 The mean, variance, and standard deviation

The mean, variance, and standard deviation of a variable that has the poisson distribution can be found by
$Mean : \mu = \lambda \; Variance : \sigma^2 = \lambda \; Standarddeviation : \sigma = \sqrt{\lambda}$

The Normal Distribution
Lecture 11

Lecture objectives

- Identify distributions as symmetric or skewed.
- Identify the properties of a normal distribution.
- Find the area under the standard normal distribution, given various z values.
- Find probabilities for a normally distributed variable by transforming it into a standard normal variable.

1 Distribution Shapes

When the data values are evenly distributed about the mean, a distribution is said to be ***a symmetric distribution.*** When the majority of the data values fall to the left or right of the mean, the distribution is said to be ***skewed.*** When the majority of the data values fall to the right of the mean, the distribution is said to be ***a negatively or left-skewed distribution.*** The mean is to the left of the median, and the mean and the median are to the left of the mode. When the majority of the data values fall to the left of the mean, a distribution is said to be ***a positively or right-skewed distribution.*** The mean falls to the right of the median, and both the mean and the median fall to the right of the mode.

2 The Normal distribution

A normal distribution is a continuous, symmetric, bell-shaped distribution of a variable
The mathematical equation for a normal distribution is

$$f(X) = \frac{e^{\frac{-(X-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

Where
$e \approx 2.718$
$\pi \approx 3.14$
$\mu = Population\ mean$
$\sigma = Population\ standarddeviation$

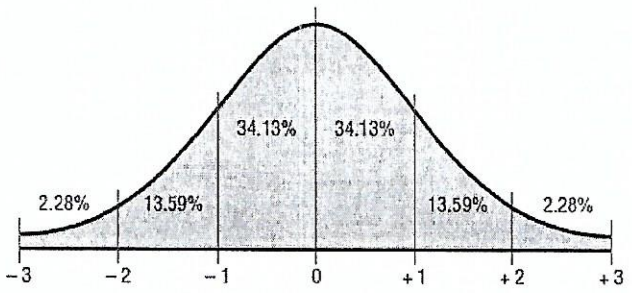
the Properties of the Theoretical Normal Distribution

1. A normal distribution curve is bell-shaped.
2. The mean, median, and mode are equal and are located at the center of the distribution.

3. A normal distribution curve is unimodal.
4. The curve is symmetric about the mean.
5. The curve is continuous. For each value of X, there is a corresponding value of Y.
6. The curve never touches the x axis. Theoretically, no matter how far in either direction the curve extends, it never meets the x axis—but it gets increasingly closer.
7. The total area under a normal distribution curve is equal to 1.00, or 100%.
8. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%.

2.1 The Standard Normal Distribution

The ***standard normal distribution*** is a normal distribution with a mean of 0 and a standard deviation of 1.



The formula for the standard normal distribution is

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

2.2 Finding Areas Under the Standard Normal Distribution Curve

First case	To the left of any z value ,Look up the z value in the table and use the area given
Second case	To the right of any z value:Look up the z value and subtract the area from 1
Third case	Between any two z values:Look up both z values and subtract the corresponding ar

Example 1. Find the area to the left of z=2.06.

Solution

First Draw the figure of the desired area.
Second We are looking for the area under the standard normal distribution to the left of z=2.06. look up the area in the table.
It is 0.9803. Hence, 98.03% of the area is less than z=2.06.

Example 2. Find the area to the right of z=-1.19.

Solution

First Draw the figure of the desired area.
Second We are looking for the area to the right of z=-1.19. This is an example of the second case. Look up the area for z=-1.19. It is 0.1170. Subtract it from 1.0000. 1.0000-0.1170=0.8830. Hence, 88.30% of the area under the standard normal distribution curve is to the right of z=-1.19

Example 3. Find the area between z=1.68 and z=-1.37.

Solution

First Draw the figure of the desired area.
Second Since the area desired is between two given z values, look up the areas corresponding to the two z values and subtract the smaller area from the larger area. The area for z=1.68 is 0.9535,and the area for z=-1.37 is 0.0853. The area between the two z values is
0.9535 -0.0853 = 0.8682 or 86.82%.

3 A Normal Distribution Curve as a Probability Distribution Curve

The area under the standard normal distribution curve can also be thought of as a probability. That is, if it were possible to select any z value at random, the probability of choosing one, say, between 0 and 2.00 would be the same as the area under the curve between 0 and 2.00

Example 4. Find the probability for each.

1. $P(0 < z < 2.32)$
2. $P(z < 1.65)$
3. $P(z > 1.91)$

Solution

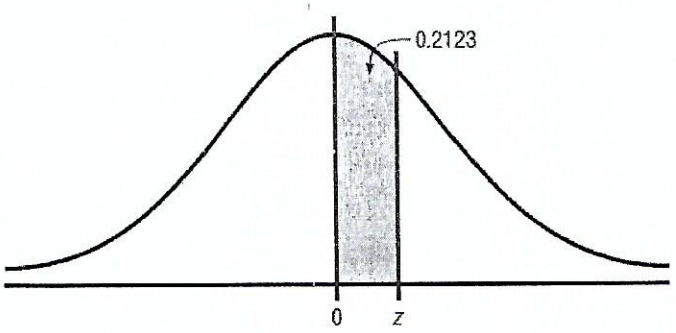
1. $P(0 < z < 2.32)$ means to find the area under the standard normal distribution curve between 0 and 2.32. First look up the area corresponding to 2.32. It is 0.9898. Then look up the area corresponding to z = 0. It is 0.500. Subtract the two areas: 0.9898 - 0.5000 = 0.4898. Hence the probability is 0.4898, or 48.98%.

2. $P(z < 1.65)$. Look up the area corresponding to z =1.65 . It is 0.9505. Hence,$P(z < 1.65) = 0.9505$, or 95.05%.

3. $P(z > 1.91)$. Look up the area that corresponds to z=1.91. It is 0.9719. Then subtract this area from 1.0000. $P(z > 1.91)$=1.0000 -0.9719 = 0.0281, or 2.81%.

Example 5. Find the z value such that the area under the standard normal distribution curve between 0 and the z value is 0.2123

Solution



In this case it is necessary to add 0.5000 to the given area of 0.2123 to get the cumulative area of 0.7123. Look up the area. The value in the left column is 0.5, and the top value is 0.06. Add these two values to get z = 0.56

4 Applications of the Normal Distribution

All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

$$z = \frac{value - mean}{standard\ deviation} \quad \text{or } z = \frac{X-\mu}{\sigma}$$

Example 6. A survey found that women spend on average \$146.21 on products during the summer months. Assume the standard deviation is \$29.44. Find the percentage of women who spend less than \$160.00. Assume the variable is normally distributed.

Solution

First	Draw the figure of the desired area.
Second	Find the z value corresponding to \$160.00
	$z = \frac{X-\mu}{\sigma} = \frac{160-146.21}{29.44} = -.47$
	Hence \$160.00 is 0.47 of a standard deviation above the mean of \$146.21
Third	Find the area, from the table.
	The area under the curve to the left of z=0.47 is 0.6808

Therefore 0.6808, or 68.08%, of the women spend less than \$160.00 on products during the summer months.

Example 7. Each month, an American household generates an average of 28 pounds of newspaper for garbage or recycling. Assume the standard deviation is 2 pounds. If a household is selected at random, find the probability of its generating

a. Between 27 and 31 pounds per month

b. More than 30.2 pounds per month

Assume the variable is approximately normally distributed.

Solution

a.

First Draw the figure and represent the area.

Second Find the two z values.

$z_1 = \frac{X-\mu}{\sigma} = \frac{27-28}{2} = 0.5$

$z_2 = \frac{X-\mu}{\sigma} = \frac{31-28}{2} = 1.5$

Third Find the appropriate area, using The normal table. The area to the left of z_2 is 0.9332, and the area to the left of z_1 is 0.3085. Hence the area between z_1 and z_2 is 0.9332- 0.3085=0.6247

Hence, the probability that a randomly selected household generates between 27 and 31 pounds of newspapers per month is 62.47%.

b.

First Draw the figure and represent the area.

Second Find the two z values.

$z = \frac{X-\mu}{\sigma} = \frac{30.2-28}{2} = 1.1$

Third Find the appropriate area. The area to the left of z=1.1 is 0.8643. Hence the area to the right of z=1.1 is 1.0000=0.8643=0.1357

Hence, the probability that a randomly selected household will accumulate more than 30.2 pounds of newspapers is 0.1357, or 13.57%.

Example 8. Americans consume an average of 1.64 cups of coffee per day. Assume the variable is approximately normally distributed with a standard deviation of 0.24 cup. If 500 individuals are selected, approximately how many will drink less than 1 cup of coffee per day?

Solution

First Draw a figure and represent the area.

Second Find the two z values.
$z = \frac{X-\mu}{\sigma} = \frac{1-1.64}{0.24} = -2.67$

Third Find the area to the left of z=-2.67. It is 0.0038

Fourth To find how many people drank less than 1 cup of coffee, multiply the sample size 500 by 0.0038 to get 1.9. the question about people, round the answer to 2 people. Hence, approximately 2 people will drink less than 1 cup of coffee a day.

TUT4

1. Determine variance and standard deviation of the number of heads in three tosses of a coin.

2. A discrete random variable X has the probability distribution given as below:

X	0	1	2	3
P(X)	k	k/ 2	k/4	K/8

1) Determine the value of k.
2) Determine $P(X \leq 2)$ and $P(X > 2)$
3) Find $P(X \leq 2) + P(X > 2)$.

3. Suppose 10,000 tickets are sold in a lottery each for Re 1. First prize is of Rs 3000 and the second prize is of Rs. 2000. There are three third prizes of Rs.500 each. If you buy one ticket, what is your expectation.